

大数据内存计算体系结构——中国需要引领性创新

中科院徐志伟

徐志伟老师的报告为一小时四十分钟，前面不到一个小时的时间里介绍了计算机发展的一些现状和老师自己的团队近年来所做的工作。之后的时间进行了相关的探讨与交流，并给出了自己对后辈的一些建议。

报告核心： 工程 → 创新 → 引领

工程即踏踏实实做研究，做实验，铺垫理论基础与实践经验;创新是在前人工作的基础上加以改进，增强已有工程;而引领则是在从未有人涉足的领域或方向，开创新的研究。

工程与创新是作为跟踪者的工作，而引领是作为创新者的工作。三者并无孰轻孰重之分，但其中任一都不可或缺，现在的中国需要加强在引领方面的建树。

中科院的三个例子

1.海云计算系统。陈氏兄弟：开辟高能效计算机新方向。2012~2016

2.内存计算体系结构。基金委重点项目。2015~2019

3.软件定义云计算。科技部重点专项。2016~2021

第三个项目如今刚刚起步，徐老师主要对前两个项目进行了讲解。

海云计算系统

挑战：能效与速度失配。计算机的发展一直稳步前进，但在近十年间，能效的提升已大大落后于速度的提升，且差距还在加大。

→ 陈氏兄弟发明寒武纪神经网络加速器。在 2014 年一年内两次分别获得 ASPLOS, Micro 两个一类会议的最佳论文。

牧本次生曲线：芯片在不断的更新换代，海云计算系统的目标是让下一代芯片性能功耗提升 1000 倍。

内存计算体系结构

算术 → 算法 → 算礼

计算机的发展从算术开始，应用的本质叫做算法，但系统的本质始终没人定义，徐老师定义其为算礼。应用的发展一直以来都非常迅速，可以很容易分析其时间复杂度，但目前来讲，我们还不能对系统

进行简洁的描述与分析。因此徐老师的团队提出了由软件定义的实用可计算系统。也就是内存计算体系结构。徐老师笑称其为“社会主义初级阶段”的算礼。

JimGray 的数据挑战：1 分钟内排序 1T 数据 → 扩展数据容量 → 新体系结构

可塑处理器 FISC, pASIC

P.S. 以上理论部分只看得懂思路，技术方面有很多没有听懂，只是知道徐老师做的这三个项目都是目前世界上没人提出过的前沿项目。

小结

1. 科学问题：现有大数据计算系统的效率很低

- 即使是 Spark 等内存计算机系统，与 HPC 相比还有 3 个数量级的提升空间
- 内存计算系统的计算机体系结构是瓶颈

2. 研究目标：提出数量级提升性能功耗比的体系结构

3. 研究思路：四类提升效率的创新技术，原型整机验证

- 可重塑处理器体系结构 FISC
- 键值对通信机制 KVC
- 用户态统一文件系统 UUFS
- 配套的编程模型与调度技术

4. 课题组已有充分的前期研究基础

- 覆盖课题主要研究内容，初步体现了创新技术点的潜力与可行性
- 10 余篇 IEEE/ACM Trans. (TOCS / TOC / TPDS /)
- 10 余篇 A 类会议 (ICDE / VLDB / MICRO /)

交流部分与我的感想

徐老师的观点是，目前中国的高校与研究院在推进中国的引领方面普遍存在一些问题。

首要问题是，当一个研究人员去申报一个新的项目的时候，上级往往会要求申请者阐述该项目的理论与实践基础。也就是说需要申请者展示前人已有的工作成果，或者外国人已经提出的相关理论，投资方也会根据项目是否有足够的理论基础来决定是否支付研究费用。长此以往，有想法的研究人员很难拿到项目与资金，这事实上就是在抹杀中国自己的引领能力与创新能力，只提倡了追着别人的脚步走，如此，中国何时能得以自强。

就如老师举的一个例子，在上世纪，“银河”发明以前，中国科学院曾经从日本购进了一台超级计算机，而自引进之日起，每一个希望使用这台计算机的用户都需要向日本申请计算机的使用权，得到日本政府的批准方可使用。中国引进计算机，实际上是支付了屈辱。

第二个问题，是中国的科研，总要求“沿途下蛋”，即科研项目最好每年甚或每半年就作出一定的阶段性成果，如果申请时，你不能说明你对项目较为具体的前瞻性成果，很可能还未开始就得不到资金与人员支持。然而事实上很多项目的研究，是十年如一日的辛苦才在一夕间实现质的蜕变。关于此事徐老师与郑老师都说过，特别是做系统的人，就因很难在短期内达到某种上级期望的成果，往往得不到支持。

如今高校，教师不拿出足数量的论文与科研成果，无以得到任何晋升，更有甚者，部分高校会根据教师论文参与会议与科学杂志的等级，赋予系数加权得到教师评分，用以衡量教师工作量的多少。这在我看来是不可思议的，科研不是盖房子，做一天工可以垒出一层砖，伟大的创新与引领整个世界进步的真正的技术少有每年都拿得出成果的。当然，我认为这种结果出现在现今以切实成绩作为唯一衡量标准的中国教育体制之下，实属正常。如果我们真的想成为世界上的科技强国，难道真的可以永远作一个老老实实的追踪者吗？我们的体制真的需要鼓励有能力有思想的人去创新。

当然这些问题的存在有它们的道理，可以一定程度的杜绝科研人员不作为，但中国的中庸之道用在科研当中，是不是太过于拖慢了中国进步的脚步？

对于这些问题，徐老师幽默的提出，我们申报项目的时候，可以“挂羊头卖狗肉”，套上一些实时最流行最打动人心的词汇，就如他所做的，做的是系统，拿的是大数据的幌子。这种上有政策下有对策的措施，让人莞然一笑的同时，令我等感到了深深的无奈。

令人欣然的是，还是有很多人看到了这些弊端，徐老师说，中科院衡量学生的标准不再是论文数量，而需要学生自我阐述做了哪些工作，由真正懂科研的人来评价其成果，如今中科院对硕士毕业已取消论文发表要求，博士生也从之篇缩减为两篇。海大计算机系主任也提出海大自身正在进行改革。

徐老师还对年轻教师的工作和本科、研究生进行了一些指导。

以老师自己为例，说明老师们想要申请重点课题，最好与其他高校或科研团体达成合作，最好能在应用层有一定的成果，且答辩之前要先完成一部分工作或者模拟过流程。而老师普遍存在的两个问题，时间不够用与杂事太多影响集中科研方面，徐老师提出这是老教师们都经历过的，建议最好每周能拿出大段的可以自己掌控的某个时间段，来集中进行科研实践。我认为对我的生活学习也有指导作用。

而对学生的要求是明白一万小时定律，即一个人在某一领域要钻研一万小时，就可以成为专家。老师提出，这一万个小时是不包括学习时间的一万小时，是动手动脑进行实践完整的一万小时。老师要求本科生最好已能完成 2000 小时，不能完成也可以完成 500 小时。我自我反省了一下总之是肯定不到的。。。徐老师说，现在中科院选择学生的标准已不再只是成绩，只问同学的最长项是什么，哪怕是概率论这门课程，只要最长板足够长，便是可塑之才。我非常同意这个观点，研究生后，发现学习模式已变成“福尔摩斯式”，你完全可以把不需要的信息全部丢开忘记，只在目前着手的这某一件案子深度扩展。

总结

首先，老师在报告中提到的软件定义的可计算系统令我很感兴趣。如果连如此复杂的系统都可以用一种应用来定义，我们的开源与操作度都会大大加强。徐老师更是提出微信可能会是下一代移动操作系统。其次便是老师对学生提出的那几个建议，我认为按照徐老师的要求我是远远难以达到的，以后应该更多的加强实践，并不断的加深自己对专一领域的理解。最后，切合本次报告的主题，我认为我们需要有这种开创性的思路，自己认为好的东西，即便是目前受到很多反对的声音，即便世界上没人提出过相似的问题，也应该有勇气去尝试。当然我目前没有这种创新的能力，先要尽力的去完成一万小时定律，积累足够的知识基础。

2016.7.26